Elsevier LaTeX template[☆]

Elsevier[1]

Radarweg 29, Amsterdam

Elsevier Inc[a,b], Global Customer Service[b,]*

[a]1600 John F Kennedy Boulevard, Philadelphia
[b]360 Park Avenue South, New York

Abstract

This template helps you to create a properly formatted LaTeX manuscript.

Keywords:

1. Main

Suppose $(X_1^T, Y_1), \ldots, (X_n^T, Y_n)$ are i.i.d. from $N_{p+1}(\mu, \Sigma)$, where $X_i \in \mathbb{R}^p$ and $Y_i \in \mathbb{R}$. Denote $X = (X_1, \ldots, X_n)$, $Y = (Y_1, \ldots, Y_n)^T$.

Write $Y = \beta_0 \mathbf{1}_n + X^T \beta + \epsilon$, where $\mathbf{1}_n$ is n dimensional vector with all elements equal to 1. ϵ has distribution $N(0, \sigma^2 I_n)$.

The problem is to test hypotheses $H : \beta = 0$.

Let $Q_n = WW^T$ be the rank decomposition of Q_n, where W_n is a $n \times n-1$ matrix with $W^T W = I_{n-1}$. The new test statistic is

$$T = \frac{y^T Q_n y}{y^T W (W^T X^T X W)^{-1} W^T y}$$

or equivalently

$$\frac{y^T Q_n y}{y^T Q_n (X^T X)^{-1} Q_n y - \left(y^T Q_n (X^T X)^{-1} \mathbf{1}_n\right)^2 / (\mathbf{1}_n (X^T X)^{-1} \mathbf{1}_n)}$$

[☆]Fully documented templates are available in the elsarticle package on CTAN.

[*]Corresponding author
Email address: `support@elsevier.com` (Global Customer Service)
URL: `www.elsevier.com` (Elsevier Inc)

[1]Since 1880.

Let $\tilde{y} = W^T y$, $\tilde{X} = XW$, $\tilde{\epsilon} = W^T \epsilon$. Then

$$\tilde{y} = \tilde{X}^T \beta + \tilde{\epsilon}$$

and

$$T = \frac{\tilde{y}^T \tilde{y}}{\tilde{y}^T (\tilde{X}^T \tilde{X})^{-1} \tilde{y}}$$

Next we derive another form of of T. We follow the similar technique of Hotelling's T^2.

Let R be an $(n-1) \times (n-1)$ orthogonal matrix satisfies

$$R\tilde{y} = \begin{pmatrix} \|\tilde{y}\| \\ 0 \\ \cdots \\ 0 \end{pmatrix}.$$

We can write

$$T = \frac{\|\tilde{y}\|^2}{\tilde{y} R^T (R\tilde{X}^T \tilde{X} R^T)^{-1} R\tilde{y}} \tag{1}$$

Denote by $B = R\tilde{X}^T \tilde{X} R^T$, then

$$T = \frac{1}{(B^{-1})_{11}}.$$

Let

$$B = \begin{pmatrix} b_{11} & b_{(1)}^T \\ b_{(1)} & B_{22} \end{pmatrix},$$

and apply the matrix inverse formula, we have $(B^{-1})_{11} = 1/(b_{11} - b_{(1)}^T B_{22}^{-1} b_{(1)})$. Hence

$$T = b_{11} - b_{(1)}^T B_{22}^{-1} b_{(1)}.$$

2. Asymptotic distribution

Note that conditioning on $\tilde{y}$, R is a constant orthogonal matrix. And $\tilde{y}$ is independent of $\tilde{X}$ under null hypotheses. So $B|\tilde{y}$ has the same distribution with

$\tilde{X}^T\tilde{X}$ under null hypotheses. Hence B is independent of $\tilde{y}$ and can be written as

$$B = \sum_{i=1}^{p} \lambda_i z_i z_i^T \tag{2}$$

where z_i's are i.i.d. $n-1$ dimensional random vectors distributed as $N(0, I_{n-1})$, $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_p > 0$ are eigenvalues of Σ_X. Denote by $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$, $Z = (Z_1, \ldots, Z_p)$. Let $Z_{(1)}$ and $Z_{(2)}$ be the first 1 row and last $n-2$ rows of Z, that is

$$Z = \begin{pmatrix} Z_{(1)} \\ Z_{(2)} \end{pmatrix}.$$

Then

$$\begin{aligned} B &= Z\Lambda Z^T \\ &= \begin{pmatrix} Z_{(1)}\Lambda Z_{(1)}^T & Z_{(1)}\Lambda Z_{(2)}^T \\ Z_{(2)}\Lambda Z_{(1)}^T & Z_{(2)}\Lambda Z_{(2)}^T \end{pmatrix}. \end{aligned} \tag{3}$$

Hence

$$\begin{aligned} T &= Z_{(1)}\Lambda Z_{(1)}^T - Z_{(1)}\Lambda Z_{(2)}^T \big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1} Z_{(2)}\Lambda Z_{(1)}^T \\ &= Z_{(1)}\big(\Lambda - \Lambda Z_{(2)}^T \big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1} Z_{(2)}\Lambda\big) Z_{(1)}^T. \end{aligned} \tag{4}$$

But

$$\begin{aligned} \mathrm{rank}(\Lambda Z_{(2)}^T \big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1} Z_{(2)}\Lambda) &= \mathrm{rank}(\Lambda^{\frac{1}{2}} Z_{(2)}^T \big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1} Z_{(2)}\Lambda^{\frac{1}{2}}) \\ &= \mathrm{rank}(I_{n-2}) = n-2, \end{aligned} \tag{5}$$

and

$$\begin{aligned} \mathrm{rank}(\Lambda - \Lambda Z_{(2)}^T \big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1} Z_{(2)}\Lambda) &= \mathrm{rank}(I_p - \Lambda^{\frac{1}{2}} Z_{(2)}^T \big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1} Z_{(2)}\Lambda^{\frac{1}{2}}) \\ &= p - n + 2. \end{aligned} \tag{6}$$

Hence

$$T \sim \sum_{i=1}^{p-n+2} \lambda_i(\Lambda - \Lambda Z_{(2)}^T \big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1} Z_{(2)}\Lambda)\chi_1^2$$

By Weyl's inequality, we have for $1 \leq i \leq p-n+2$

$$\lambda_i(\Lambda - \Lambda Z_{(2)}^T \big(Z_{(2)}\Lambda Z_{(2)}^T\big)^{-1} Z_{(2)}\Lambda) \leq \lambda_i(\Lambda), \tag{7}$$

and

$$
\begin{aligned}
&\lambda_i(\Lambda - \Lambda Z_{(2)}^T (Z_{(2)}\Lambda Z_{(2)}^T)^{-1} Z_{(2)}\Lambda) \\
\geq &\lambda_{i+n-2}(\Lambda) + \lambda_{p-n+2}(-\Lambda Z_{(2)}^T (Z_{(2)}\Lambda Z_{(2)}^T)^{-1} Z_{(2)}\Lambda) \\
= &\lambda_{i+n-2}.
\end{aligned} \tag{8}
$$

Hence

$$\sum_{i=n-1}^{p} \lambda_i \chi_1^2 \leq T \leq \sum_{i=1}^{p-n+2} \lambda_i \chi_1^2$$

Note that under condition $\mathrm{tr}\Sigma^4/(\mathrm{tr}\Sigma^2)^2 \to 0$, we have by Liapounoff central limit theorem that

$$\frac{\sum_{i=1}^{p} \lambda_i \chi_1^2 - \mathrm{tr}\Sigma_X}{\sqrt{\mathrm{tr}(\Sigma_X^2)}} \xrightarrow{\mathcal{L}} N(0,1).$$

And

$$\frac{T - \mathrm{tr}\Sigma_X}{\sqrt{\mathrm{tr}(\Sigma_X^2)}} - \frac{\sum_{i=1}^{p} \lambda_i \chi_1^2 - \mathrm{tr}\Sigma_X}{\sqrt{\mathrm{tr}(\Sigma_X^2)}} = \frac{T - \sum_{i=1}^{p} \lambda_i \chi_1^2}{\sqrt{\mathrm{tr}(\Sigma_X^2)}}, \tag{9}$$

To prove (9) $\xrightarrow{P} 0$, we only need to prove

$$\mathrm{E}\Big(\frac{\sum_{i=1}^{n-2} \lambda_i \chi_1^2}{\sqrt{\mathrm{tr}(\Sigma_X^2)}}\Big) \to 0,$$

that is

$$\mathrm{E}\Big(\frac{\sum_{i=1}^{n-2} \lambda_i}{\sqrt{\mathrm{tr}(\Sigma_X^2)}}\Big) \to 0. \tag{10}$$

If λ_i's are bounded below and above, then (10) is equivalent to

$$n/\sqrt{p} \to 0, \tag{11}$$

or $p/n^2 \to \infty$. We thus obtain the following theorem.

Theorem 1. *Suppose*

$$E\Big(\frac{\sum_{i=1}^{n-2} \lambda_i}{\sqrt{tr(\Sigma_X^2)}}\Big) \to 0,$$

and

$$\frac{tr\Sigma^4}{(tr\Sigma^2)^2} \to 0.$$

Then under null hypotheses, we have

$$\frac{T - tr\Sigma_X}{\sqrt{tr(\Sigma_X^2)}} \xrightarrow{\mathcal{L}} N(0,1).$$

3. Full Asymptotic Results

$\Sigma_X = P\Lambda P^T$

Non-spike: there's no principal component ($r = 0$). That is, $\lambda_1 = \cdots = \lambda_p$.

Spike: there's r principal components. That is, $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_r \geq \lambda_{r+1} = \cdots = \lambda_p$. Denote by P_1 the first r column of P and P_2 the last $p - r$ column of P.

$$\begin{aligned} Y &= \beta_0 \mathbf{1}_n + X^T\beta + \epsilon \\ &= \beta_0 \mathbf{1}_n + X^T P_1 P_1^T \beta + X^T P_2 P_2^T \beta + \epsilon \end{aligned} \tag{12}$$

In either case, let λ be $\lambda = \lambda_{r+1} = \cdots = \lambda_p$.

PCR try to do regression between Y and (estimated) $X^T P_1$. If P_1 is observed, then the problem is reduced to testing an ordinary regression model. However, it's not the case.

Simply estimating P_1 and invoke classical testing procedure may not be a good idea since the estimation may not be consistent in high dimension. In fact, there may be even no principal component!

In this paper, testing PCR means testing:

H_0: There's no principal component or there's r principal components but $P_1^T\beta = 0$.

H_1; There's r principal components and $P_1^T\beta \neq 0$.

Next we consider:

1. There's no PC.
2. There's r principal components but $P_1\beta = 0$.

3.1. circumstance 1

$$T = \frac{\beta^T \tilde{X}\tilde{X}^T\beta + 2\beta^T\tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{\beta^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\beta + 2\beta^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \tag{13}$$

Independent of data, generate a random p dimensional orthonormal matrix O with Haar invariant distribution. And

$$T = \frac{(O\beta)^T O\tilde{X}(O\tilde{X})^T O\beta + 2(O\beta)^T \tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{(O\beta)^T O\tilde{X}((O\tilde{X})^T O\tilde{X})^{-1}(O\tilde{X})^T \beta + 2(O\beta)^T O\tilde{X}((O\tilde{X})^T O\tilde{X})^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T((O\tilde{X})^T O\tilde{X})^{-1}\tilde{\epsilon}} \tag{14}$$

Note that conditioning on O, $O\tilde{X}$ is a random matrix with each entry independently distributed as $N(0,\lambda)$. Hence O is independent of $O\tilde{X}$. Observe also that $O\beta/\|\beta\|$ is uniformly distributed on the unit ball. We can without loss of generality assuming that $\beta/\|\beta\|$ is uniformly distributed on the unit ball. Independent of data, generate $R > 0$ with R^2 distributed as χ^2_p. Then $\xi = R\beta/\|\beta\|$ distributed as $N_p(0, I_p)$. Note that conditioning on $\tilde{X}$, $\eta = (\tilde{X}^T\tilde{X})^{-1/2}\tilde{X}^T\xi$ is distributed as $N_{n-1}(0, I_{n-1})$. Hence η is independent of $\tilde{X}$.

Then

$$\begin{aligned} T &= \frac{(\|\beta\|/R)^2\xi^T\tilde{X}\tilde{X}^T\xi + 2(\|\beta\|/R)\xi^T\tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{(\|\beta\|/R)^2\xi^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\xi + 2(\|\beta\|/R)\xi^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \\ &= \frac{(\|\beta\|/R)^2\eta^T\tilde{X}^T\tilde{X}\eta + 2(\|\beta\|/R)\eta^T(\tilde{X}^T\tilde{X})^{1/2}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{(\|\beta\|/R)^2\eta^T\eta + 2(\|\beta\|/R)\eta^T(\tilde{X}^T\tilde{X})^{-1/2}\tilde{\epsilon} + \tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \end{aligned} \tag{15}$$

We consider

$$\frac{\frac{\|\beta\|^2}{R^2}\eta^T\eta}{\tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} = \frac{(n-1)\|\beta\|^2}{p}\frac{\frac{p}{R^2}\frac{\eta^T\eta}{n-1}}{\tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \tag{16}$$

Note that

$$\begin{aligned} \tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon} &\sim \frac{\tilde{\epsilon}^T\tilde{\epsilon}}{(\tilde{X}^T\tilde{X})_{11\cdot 2}} \\ &\sim \frac{\sigma^2}{\lambda}\frac{\chi^2_{n-1}}{\chi^2_{p-n+2}} \end{aligned} \tag{17}$$

Hence

$$\begin{aligned} \frac{\frac{\|\beta\|^2}{R^2}\eta^T\eta}{\tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} &= \frac{(n-1)(p-n+2)}{p(n-1)}\frac{\lambda}{\sigma^2}\|\beta\|^2(1+o_P(1)) \\ &= \frac{\lambda}{\sigma^2}\|\beta\|^2(1+o_P(1)) \end{aligned} \tag{18}$$

On the other hand,

$$\frac{\frac{\|\beta\|^2}{R^2}\eta^T\tilde{X}^T\tilde{X}\eta}{\tilde{\epsilon}^T\tilde{\epsilon}} = \frac{\|\beta\|^2}{\sigma^2 p(n-1)}\frac{\frac{p}{R^2}\eta^T\tilde{X}^T\tilde{X}\eta}{\frac{\tilde{\epsilon}^T\tilde{\epsilon}}{\sigma^2(n-1)}} \tag{19}$$

Note that

$$\begin{aligned} \eta^T\tilde{X}^T\tilde{X}\eta &\sim \eta^T\eta(\tilde{X}^T\tilde{X})_{11} \\ &\sim \lambda\chi^2_{n-1}\chi^2_p \end{aligned} \tag{20}$$

Hence

$$\begin{aligned} \frac{\frac{\|\beta\|^2}{R^2}\eta^T\tilde{X}^T\tilde{X}\eta}{\tilde{\epsilon}^T\tilde{\epsilon}} &= \frac{\|\beta\|^2}{\sigma^2 p(n-1)}\lambda(n-1)p(1+o_P(1)) \\ &= \frac{\lambda}{\sigma^2}\|\beta\|^2(1+o_P(1)) \end{aligned} \tag{21}$$

We can deduce that: If $\|\beta\|^2 \to 0$, then

$$T = \frac{\tilde{\epsilon}^T\tilde{\epsilon}}{\tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}}(1+o_P(1)) \tag{22}$$

If $\|\beta\|^2 \to \infty$, then

$$T = \frac{\eta^T\tilde{X}^T\tilde{X}\eta}{\eta^T\eta}(1+o_P(1)) \tag{23}$$

3.2. circumstance 2

$$T = \frac{\beta^T\tilde{X}\tilde{X}^T\beta + 2\beta^T\tilde{X}\tilde{\epsilon} + \tilde{\epsilon}^T\tilde{\epsilon}}{\beta^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\beta + 2\beta^T\tilde{X}(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon} + \tilde{\epsilon}^T(\tilde{X}^T\tilde{X})^{-1}\tilde{\epsilon}} \tag{24}$$

4. Simulation Results

References